

DOI:10.19651/j.cnki.emt.2107978

基于全局多粒度池化的可见光红外行人重识别^{*}

周航 黄春光 程海

(黑龙江大学电子工程学院 哈尔滨 150000)

摘要: 可见光红外行人重新识别是一种跨模态检索的问题。由于可见光和红外图像模态差异较大,能够精确的匹配行人仍然具有很大的挑战。最近的研究表明,利用池化描述身体部位的局部特征以及人图像本身的全局特征,即使在身体部位缺失的情况下,也能给出鲁棒的特征表示,但是简单的全局平均池化很难获取行人的细节特征。针对这个问题,提出一种新的全局多粒度池化的方法,利用全局平均池化和全局最大池化结合的方法,提取行人更多的背景和纹理信息。此外,传统的三元组损失在跨模态行人重识别上效果并不好。设计了一种新的跨模态三元损失,以优化类内和类间距离,并监督网络学习有区别的特征表示。通过实验证明了所提方法的有效性,并在 RegDB 和 SYSU-MM01 数据集上分别取得了 88.01% Rank-1, 79.26% mAP, 和 60.24% Rank-1, 57.50% mAP 的结果。

关键词: 全局多粒度池化;可见光红外行人重识别;困难跨模态三元损失

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Visible infrared person re-identification based on global multi-granularity pooling

Zhou Hang Huang Chunguang Cheng Hai

(School of Electronic Engineering Institute, Heilongjiang University, Harbin 150000, China)

Abstract: Visible infrared person re-identification is a cross-modal retrieval problem. Being able to accurately match pedestrians remains challenging due to the large modal differences between visible and infrared images. Recent research has shown that using pooling to describe local features of body parts as well as global features of the human image itself can give a robust feature representation even when body parts are missing, but simple global average pooling is difficult to obtain detailed features of pedestrians. To address this problem, this paper proposes a new global multi-granularity pooling approach that uses a combination of global average pooling (GAP) and global maximum pooling (GMP) to extract more background and texture information of person. In addition, the traditional triplet loss does not work well for cross-modal person re-identification. We design a new cross-modal triplet loss to optimise intra-class and inter-class distances and supervise the network to learn differentiated feature representations. We experimentally demonstrate the effectiveness of the proposed method and achieves 88.01% Rank-1, 79.26% mAP, and 60.24% Rank-1, 57.50% mAP on the RegDB and SYSU-MM01 datasets, respectively.

Keywords: global multi-granularity pooling; visible thermal person re-identification; hard cross-modality triplet loss

0 引言

行人重识别^[1]能够在不相交的多个摄像机之间匹配同一个人。在实际生活中应用性很强,由此得到了学术界的广泛关注。先前的工作主要关注于可见光的行人重识别问题(即探针图像和图库图像集都是可见光图像),并取得了很好的效果。但在光线不好的情况下,可见光摄像头不能拍到清晰的行人图像,以至于传统的行人重识别模型无法

达到理想的效果。与之相比,红外摄像机对光照条件并不敏感,在弱光下均可成像。所以,利用可见光和红外图像的行人重识别逐渐变成了新的研究热点。

可见光红外行人重识别,顾名思义指的是可见光图像和红外图像之间的相互匹配。由于红外图像^[2-3]不同于可见光图像,为单通道图像,且两者的波长范围也大不相同。此外,由于视角的改变,摄像机获取到的人物图像通常遭受很明显的类内变化,如姿态、背景和是否被遮挡的变化。所

收稿日期:2021-09-28

^{*} 基金项目:国家自然科学基金青年基金(51607059)、黑龙江大学基础科学研究项目(KJCX201904,2020-KYYWF-1001)资助

以,很难得到精确的匹配,给可见光红外行人重识别带来了极大的挑战。为了解决这些问题,Wu等^[4]设计了一个零填充网络,并提出跨模态行人重识别公开数据集 SYSU-MM01。Ye等^[5]提出一个双向双约束的顶级损失来学习判别特征表示。与此同时,Dai等^[6]使用了1个用于学习图像表示的生成器和1个用于识别RGB和IR图像的模式鉴别器来学习两种模态之间的区别。随后,Zhang等^[7]利用模态融合的方法,在不丢失空间结构的情况下将交叉模态图像嵌入到一个普通的三维张量空间中。然而,上述这些方法,都是采用单一全局分量,并没有考虑局部和全局结合的方法。此外,传统的提取全局特征的方法均采用GAP的方法,忽略了行人的细节信息造成识别不准确等问题。

针对这些问题,提出全局多粒度池化可见光红外行人重识别网络(GMP-VTReID)。不同于传统的two-stream结构,本文采用经过ImageNet预训练的ResNet50作为基础网络,并将其后3层进行权重共享,以便更好地利用两种模态的信息。此外,全局和局部分支被使用获取不同层次上的行人特征,以增强模型的鲁棒性。在全局部分中,全粒度池化被提出,通过GAP和GMP跳级连接的方法得到更多的行人背景和纹理信息。而在局部部分中,采用PCB^[8]

的方法,将特征水平分割成6个部分。同时,希望两种模态减少类内变化增大类间变化。本文设计了一种新的跨模态三元损失,并在公开的数据集上取得了很好的结果。

1 模型结构

1.1 整体框架

GMP-VTReID的整体框架如图1所示,由于网络的输入包括可见光和红外图像,且红外图像不同于可见光图像,没有色彩纹理等特性,而传统的one-stream网络很难学习到每个模态的特征。因此,采用two-stream结构, I_V 表示可见光图像, I_T 代表近红外图像。简单来说,采用经过预训练后的ResNet50作为基础网络结构用函数 φ 表示。此外,为了扩大接收场的面积,丰富特征粒度,GMP-VTReID删除了ResNet50最后的下采样层。随后,GMP-VTReID分为两个分支:全局分支与局部分支。在局部分支中,延续水平分割的方法,将得到的特征均匀分成6份以增强模型的鲁棒性。在全局分支中,则采用先将基础网络提取到的两种特征进行融合得到 f_g 如式(1)所示,再利用GAP和GMP级联的方法获得全粒度池化模块,对行人的背景和纹理信息进行提取。

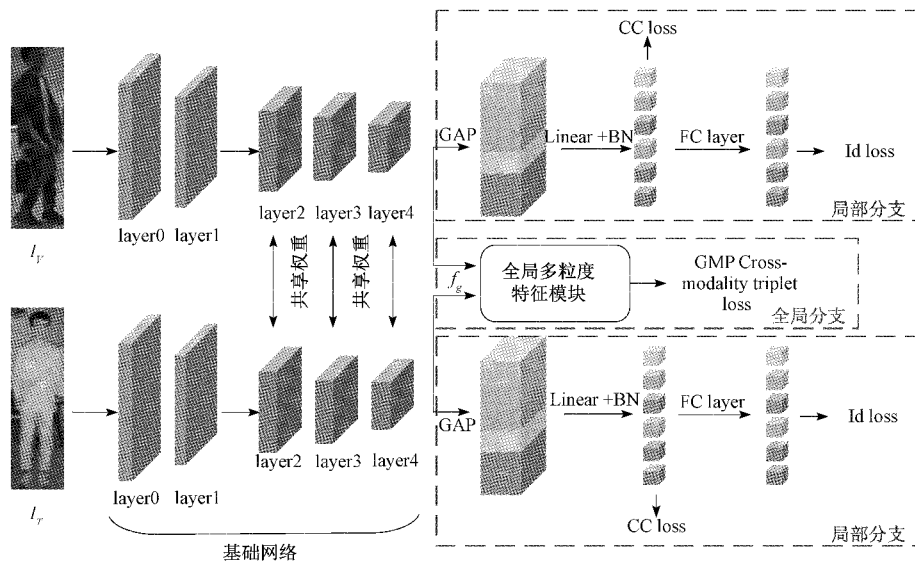


图1 总框架图

$$f_g = T(\varphi(I_V), \varphi(I_T)) \quad (1)$$

式中: T 为concatenate操作,这里表示按行拼接。

1.2 全局多粒度特征模块

目前,GAP和GMP是两种最常见的提取全局特征的方法可见光红外行人重识别中。如图2所示,一般来说,GAP是每个空间位置具有用于“期望特征”的检测器,平均每个空间位置,其行为类似于平均输入图像的不同平移的预测,所以GAP可以更多地保留图像的背景信息。相反,GMP对邻域内特征点取最大值,GMP所能够更多注重行人的纹理信息。单独使用其中某一个都会造成特征的部

分缺失。为了能够兼容背景和纹理信息,提出全局多粒度池化模块,利用两种池化相减的方法(GAP-GMP)可以获取更多除行人以外的背景信息。

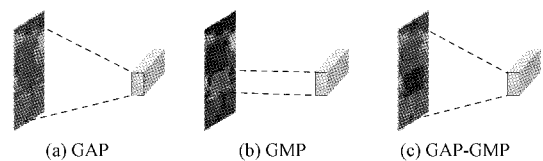


图2 不同池化的感受野

具体来讲如图3所示,已知近红外和可见光融合后的

特征 f_g , 分别经过 GMP 和 GAP 提取后得到 f_{gm} 和 f_{ga} 。采取 GAP-GMP 的方法, 获得更多的行人对比特征。随后, 将两个特征分别送入由 Linear 和 Batch Normalization 构成的 Fc layer 将每一部分的特征降维, 由 2 048 维的特征降至 512 维, 分别得到 f'_{gm} 和 f'_{ga} 。本文希望最终获取到的特征既包含最有区别的细节特征, 又包含行人对比特征。因此, 全粒度池化模块采取跳级连接的方法(增强行人的可分辨性), 得到全局特征 g , 如式(2)所示。

$$g = f'_{gm} + L(T(f'_{gm}, f'_{ga})) \quad (2)$$

式中: L 为 FC 层, T 为 concatenation 操作, 这里表示按列拼接。

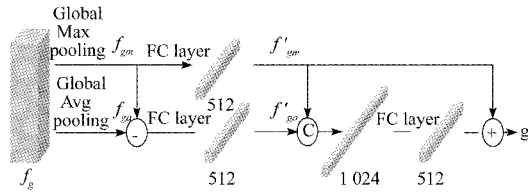


图 3 全局多粒度池化模块结构

1.3 损失函数

1)GMP cross-modality triplet loss

在基于困难三元组损失(Hard Triplet loss)的网络中, 输入通常是一系列元组, 其中 X^a 代表锚(Anchor)图像, X^p 和 X^n 分别代表正样本和负样本图像。它的意义是为了保证 Anchor 图像更接近正样本图像而远离负样本图像。假设选择 N 个元组进行训练, 其困难三元组损失函数可以定义如下:

$$L_{lri} = \sum_{i=1}^N [\max \|f(X_i^a) - f(X_i^p)\|_2^2 - \min \|f(X_i^a) - f(X_i^n)\|_2^2 + \alpha]_+ \quad (3)$$

式中: $[b]_+ = \max(b, 0)$, α 是人工定义的正样本和负样本之间的最小距离。

将 Hard Triplet loss 直接使用在跨模态行人重识别的场景中, 会造成相同 id 的两种模态距离过远, 以至于无法区别两种模态是否来自同一个人, 如图 4(a) 所示。因此, 本文受 Hard Triplet loss^[9] 启发的, 提出了一种新的 GMP cross-modality triplet loss 如图 4(b) 所示。主要由交叉模态和类内模态损失构成。其中, 交叉模态损失 L_{gmp-c} 如式(4)所示, 其目的是缩小同一个人 id 之间的跨模态差异。

$$L_{gmp-c} = \sum_{i=1}^N [\max \|f(V_i^a) - f(T_i^p)\|_2^2 - \min \|f(V_i^a) - f(T_i^n)\|_2^2 + \beta]_+ \quad (4)$$

此外, 式(5)中为了解决不同视角、姿势和遮挡等模态类内的变化, GMP 损失引入类内模态约束 L_{gmp-i} 。该约束旨在让类内跨模态变量应小于类间相同模态变量, 以达到消除图像补丁层上的模态差距的目的。

$$L_{gmp-i} = \sum_{i=1}^N [\max \|f(V_i^a) - f(T_i^p)\|_2^2 - \min \|f(V_i^a) - f(V_i^n)\|_2^2 + \beta]_+ \quad (5)$$

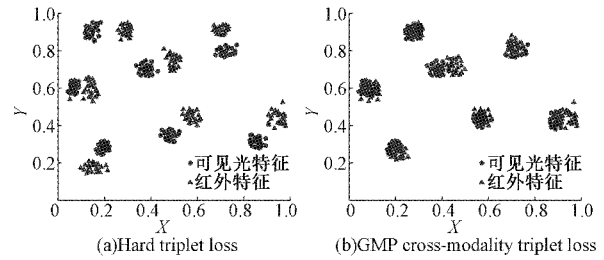


图 4 两种损失实现模型的特征分布(该特征来自于利用两种损失训练的模型, 提取 SYSU-MM01 数据集中 7 个身份的 46 张图片, 并通过 t-SNE 方法将特征降至 2 维)

综上所述, GMP cross-modality triplet loss 由交叉模态困难三元损失和类内模态困难三元损失。其公式如式(6)所示。

$$L_{gmp} = L_{gmp-c} + L_{gmp-i} \quad (6)$$

2)Cross-modality center loss

以前的方法, 都只是在于利用两种模态进行损失的计算, 但是由于传统的损失是利用每一个特征进行计算的。如果在这些特征中存在一些异常值, 可能造成无法约束两个特征的距离, 这也会导致形成错误的三元组值, 对后面的训练造成影响。针对此问题, 式(7)和(8)中, 本文使用 Cross-modality center loss(CC loss)^[10] 来缩小同一个人的两种模态间的差距。并通过监督训练很好地扩大类间特征的差异, 提升类内跨模态相似度。

$$\begin{cases} C_V = \frac{1}{M} \sum_{j=1}^M V_j \\ C_T = \frac{1}{N} \sum_{j=1}^N T_j \end{cases} \quad (7)$$

$$L_{CC} = \sum_{i=1}^K [\|C_V^i - C_T^i\|_2^2 + \rho]_+ \quad (8)$$

式中: C_V, C_T 分别代表可见光和红外图像的聚类中心, K 表示分类总数, $\|A - B\|_2^2$ 为欧氏距离的平方, ρ 是人工定义两种聚类中心之间的最小距离。

此外, 同一个人即使在不同姿势、视角、光照等变化的情况下, 两种模态的特征也应该具有较高的相似度。因此本文引入身份损失(id loss):

$$L_{id} = -\frac{1}{n} \sum_{i=1}^n \log \frac{e^{y_i}}{\sum_{j=1}^K e^{y_j}} \quad (9)$$

式中: n 为图片总数, y_j 是每个局部特征经过全连接层后的输出, l_i 是第 i 张图片相对应的标签。

综上所述, GMP-VTRcID 的整体训练损失如下:

$$L = L_{id} + \theta L_{CC} + \lambda L_{gmp} \quad (10)$$

式中: θ 是 CC loss 的权重, λ 为 GMP cross-modality triplet loss 的权重。

2 实验与分析

在本节中将在两个公开的数据集(SYSU-MM01^[11]和

RegDB^[11]) 上评估 GMP-VTReID 框架, 验证 GMP-VTReID 框架的有效性。

2.1 数据集和实施细节

本文使用两种跨模态行人重识别数据集, 即 SYSU-MM01^[4] 和 RegDB^[11] 上评估了所提出的方法。其中 SYSU-MM01 数据集, 包括来自 4 个可见光摄像机和 2 个红外摄像机的 491 个行人。训练集包含 19 659 幅可见图像和 395 人的 12 792 幅红外图像, 测试集包含 96 人。有两种测试模式, 即全搜索模式和室内搜索模式。本文从这两个方面对系统进行评估。RegDB 是由 1 个可见光和 1 个红外摄像机收集的。共有 412 人, 每个人有 10 个可见光图像和 10 个热图像。按照评估协议^[12] 将数据集随机分成两半, 分别用于训练和测试。每个数据集测试均采用两种搜索方式(红外图像查找可见光图像、可见光图像搜索红外图像)。并且, GMP-VTReID 利用 Pytorch 框架将 $288 \times 144 \times 3$ 的行人图像在 Intel(262v4@2.10 GHz CPU, GeForce GTX 2080 Ti GPU 和 128 G RAM) 上训练 60 代。同时训练 batchsize 为 32, 验证 batchsize 为 64。此外, 本文采用 SGD 优化器对网络进行优化, 动量参数设为 0.9, ρ 设置为 0.1, β 设为 0.3, θ 取 0.5。

为了评价本文提出的方法, 使用累积匹配特征(CMC)和平均精度(mAP)作为评价指标。CMC 曲线是算一种 top-k 的击中概率, 主要用来评估闭集中 Rank 的正确率。

而 mAP 是用来衡量检索性能, 主要用于多目标检测和多标签图像分类模型中, 即将单一分类任务中的平均精度 AP (average precision) 求和再取平均得到的。

2.2 与先进方法进行对比

为了更好地证明 GMP-VTReID 框架的有效性, 本文分别在 SYSU-MM01 和 RegDB 两个数据集, 且从两个搜索方式与现有的方法进行对比, 如表 1、2 所示。本文选取了近几年发表高水平会议和期刊上的论文的实验结果进行对比。GMP-VTReID 代表本文所提出的算法。从表中可以看出, 相对于传统的 one-stream 网络 Zero-padding^[4] 逐渐被能够更好地提取每一种模态的 two-stream 网络 EDFL^[16]、MACE^[18]、DDAG^[19] 所代替。此外, 在 RegDB 数据集上, GMP-VTReID 的方法相比 DGTL^[21] 在两种搜索方式下最大提升 Rank-1 为 2.9, mAP 为 2.37。对于 EDFL^[16] 来讲, 本文的方法能够增加 35.43 的 Rank-1 和 26.22 的 mAP。而 SYSU-MM01 数据集, 本文提出的方法比 TSLFN+HC^[10] 单纯的水平分割的方法, 在 Rank-1 上提升 3.28, mAP 增长 2.55 个点。证明了 two-stream 提取网络且局部特征和全局特征结合的方法, 模型更具有鲁棒性。具体来说, 本文的框架在 RegDB 数据集上获得了 88.01% 的 Rank-1 和 79.26% 的 mAP, 在 SYSU-MM01 数据集上达到 60.24% 的 Rank-1 和 57.50% 的 mAP。

表 1 在 RegDB 数据上与其他方法的比较

方法	可见光检索红外				红外检索可见光			
	Rank-1	Rank-10	Rank-20	mAP	Rank-1	Rank-10	Rank-20	mAP
Zero-padding ^[4]	17.75	56.42	67.52	31.83	16.63	34.68	44.25	17.82
HCML ^[12]	24.44	47.53	56.78	20.80	21.70	45.02	55.58	22.24
HSME ^[13]	41.34	65.21	75.13	38.82	40.67	65.35	75.27	37.50
D2RL ^[14]	43.40	66.10	76.30	44.10	—	—	—	—
AlignGAN ^[15]	57.90	—	—	53.60	56.30	—	—	53.40
EDFL ^[16]	52.58	72.10	81.47	52.98	51.89	72.09	81.04	52.13
XIVReID ^[17]	62.21	83.13	91.72	60.18	—	—	—	—
MACE ^[18]	72.37	88.40	93.59	69.09	72.12	88.07	93.07	68.57
DDAG ^[19]	69.34	86.19	91.49	63.46	68.06	85.15	90.31	61.80
HMAN ^[20]	83.45	94.66	96.65	80.01	82.72	93.69	96.65	77.82
GMP-VTReID	88.01	96.02	98.06	79.26	87.17	95.93	97.76	78.45

此外, 本文为了更全面地证明 GMP-VTReID 模型的有效性。对于 SYSU-MM01 数据集增添了多搜索的搜索模式, 如表 3、4 所示。对于多次全搜索模式, 本文所提出的方法相比于 TSLFN+HC^[10], Rank-1 从 62.09% 到 66.11%, mAP 从 48.02% 到 51.03%。GMP-VTReID 在室内搜索的情况下全搜索模式, 本文所提出的方法相比于 TSLFN+HC^[10], Rank-1 从 62.09% 到 66.11%, mAP 从

48.02% 到 51.03%。GMP-VTReID 在室内搜索的情况下 Rank-1 和 mAP 略低于 HMAN^[20]。但是, 在 Rank-10 和 Rank-20 的精度高于 HMAN^[20]。综上所述, GMP-VTReID 的优势主要是在于采用双流的基础网络分为局部和全局分支, 从不同的角度提取更细节的特征。同样的, GMP-VTReID 在其他设置中也可以观察到类似的结果。证明了本文提出的网络的性能更为成功。

表 2 在 SYSU-MM01 数据上与其他方法的比较

方法	全搜索(单搜索)				室内搜索(单搜索)			
	Rank-1	Rank-10	Rank-20	mAP	Rank-1	Rank-10	Rank-20	mAP
Zero-padding ^[4]	14.80	54.12	71.33	15.95	20.58	68.38	85.79	26.92
HCML ^[12]	14.32	53.16	69.17	16.16	24.52	73.25	86.73	30.08
HSME ^[13]	20.68	62.74	77.95	23.12	—	—	—	—
D2RL ^[14]	28.90	70.60	82.40	29.20	—	—	—	—
AlignGAN ^[15]	42.40	85.00	93.70	40.70	45.90	87.60	94.40	54.30
EDFL ^[16]	36.94	85.42	93.22	40.77	—	—	—	—
XIVReID ^[17]	49.92	89.79	95.96	50.73	—	—	—	—
MACE ^[18]	51.64	87.25	94.44	50.11	57.35	93.02	97.47	64.79
DDAG ^[19]	54.75	90.39	95.81	53.02	61.02	94.06	98.41	67.98
TSLFN+HC ^[10]	56.96	91.50	96.82	54.95	59.74	92.07	96.22	64.91
DGTL ^[21]	57.34	—	—	55.13	63.11	—	—	69.20
GMP-VTReID	60.24	92.95	98.08	57.50	61.50	96.88	99.46	68.64

表 3 在 SYSU-MM01 数据集上与最先进的作品比较

方法	全搜索(多搜索)			
	Rank-1	Rank-10	Rank-20	mAP
Zero-padding ^[4]	19.13	61.40	78.41	10.89
JSIA ^[22]	45.10	85.70	93.80	29.50
DGD+MSR ^[23]	43.86	86.94	95.68	30.48
TSLFN+HC ^[10]	62.09	93.74	97.85	48.02
HMAN ^[20]	59.10	87.40	93.11	50.08
GMP-VTReID	66.11	94.93	98.03	51.03

表 4 在 SYSU-MM01 数据集上与最先进的作品比较

方法	室内搜索(多搜索)			
	Rank-1	Rank-10	Rank-20	mAP
Zero-padding ^[4]	24.43	75.86	91.32	18.64
JSIA ^[22]	52.70	91.10	96.40	42.70
DGD+MSR ^[23]	46.56	93.57	98.80	40.08
TSLFN+HC ^[10]	69.76	95.85	98.90	57.81
HMAN ^[20]	76.43	93.83	96.84	68.43
GMP-VTReID	75.18	98.28	99.59	62.70

2.3 提出方法的消融研究

为了证实本文提出的全粒度池化模块的有效性,从两个方面对比了是否添加该模块对系统识别精度的影响,如表 5、图 5 所示。表 5 为两种情况下的网络模型识别精度, baseline 代表没有添加全粒度模块的方法, GMP 表示已经添加过的方法。从表 5 中可以看出,在两个数据集上 GMP

的 Rank-1 和 mAP 均得到了大幅度的提升。此外,图 5 为有无添加该模块对训练时精度的影响, w/o 表示添加过该模块的方法, w 则是没有添加时的情况。在图 5 中明显可以看到,添加全粒度池化模块的精度曲线高于未添加的情况。综上所述,证明添加全粒度池化模块对于可见光红外行人重识别模型的效果得到提升,增强模型的鲁棒性。

表 5 与基础网络的比较

方法	RegDB 数据集				SYSU-MM01 数据集			
	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
Baseline	86.55	95.61	97.79	76.39	58.66	90.63	97.54	55.42
GMP	88.01	96.02	98.06	79.26	60.24	92.95	98.08	57.50

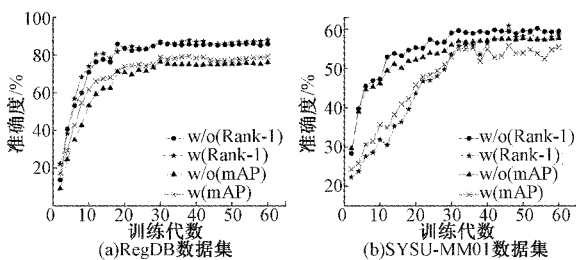


图 5 训练曲线图

2.4 参数取值分析

图 6 所示为 GMP cross-modality triplet loss (L_{gmp}) 中的权重 λ 取值对识别性能的影响。选取 λ 从 0~1.0, 每隔 0.1 取一个点, 记录此时的系统精度。从图 6 中可以得到, 随着 λ 的逐渐增加, 模型的精度也逐渐增加, 当 λ 达到 1 时系统性能达到顶峰。图 7 则代表 Cross-modality center loss (CC loss) 的权重 θ 取值对识别精度的影响。具体的, θ 选取 5 个点, 0.1、0.3、0.5、0.7、0.9, 记录每个值得

到的精度。从图7中可以看出, θ 在 0~0.5 区间时精度逐渐提升。但是在 0.5 时达到顶峰, 后续逐渐下降。主要是因为, 裕度参数被分配一个较大或较小的值时, 该模型将导致较弱的性能。这可能是由于太大的值会使网络难以优化, 太小的值会使模型过拟合。总结来说, 当 θ 为 0.5 和 λ 为 1 时, 模型分别在 RegDB 和 SYSU-MM01 数据集上实现了最佳性能。

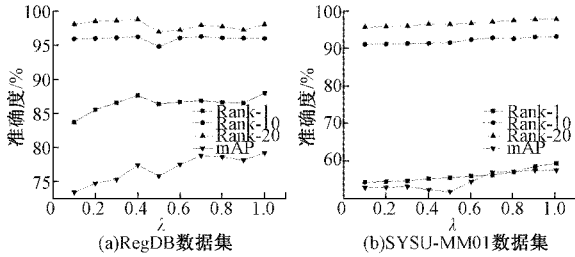


图6 不同 λ 得到的精度曲线

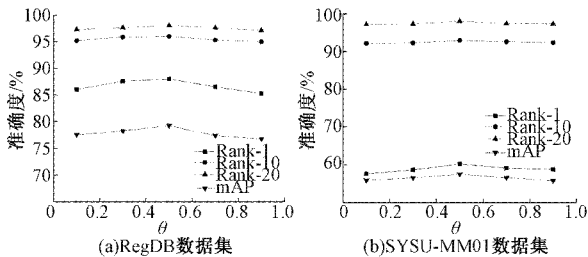


图7 不同 θ 得到的精度曲线

3 结 论

本文提出了一种新的全局多粒度池化可见光红外行人重识别网络。它使用双流的网络结构通过全局和部分特征结合的方法。提出了全粒度池化方法, 利用全局平均池化和全局最大池化通过相减和跳级连接的模式, 获取更多的行人背景和细节信息, 增大模型的鲁棒性。此外, 本文设计了一种新的跨模态三元损失函数, 解决了传统跨模态三元损失无法挖掘类内距离最大化和类间距离最小化的三元组, 导致网络模型识别效果不佳的问题。在两个公开的数据集上进行了实验, 实验结果证明了本文提出的模型对可见光红外行人重识别的有效性。

参考文献

- [1] 刘仁春, 孟朝晖. 基于孪生对抗 SGAN 的行人重识别研究[J]. 电子测量技术, 2019, 42(15): 155-160, 166.
- [2] 李春艳, 孙韬, 谢俊峰. EMF 深度学习可见光/红外图像融合算法[J]. 国外电子测量技术, 2020, 39(10): 25-32.
- [3] 姜骞, 刘亚东, 方健, 等. 基于轮廓特征的电力设备红外和可见光图像配准方法[J]. 仪器仪表学报, 2020, 41(11): 252-260.
- [4] WU A, ZHENG W S, YU H X, et al. RGB-infrared

cross-modality person re-identification[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 5380-5389.

- [5] YE M, WANG Z, LAN X, et al. Visible thermal person re-identification via dual-constrained top-ranking[C]. IJCAI, 2018: 1092-1099.
- [6] DAI P, JI R, WANG H, et al. Cross-modality person re-identification with generative adversarial training [C]. IJCAI, 2018: 677-683.
- [7] ZHANG S, YANG Y, WANG P, et al. Attend to the difference: Cross-modality person re-identification via contrastive correlation [J]. ArXiv Preprint, ArXiv:1910.11656.
- [8] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 480-496.
- [9] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J]. ArXiv Preprint, 2017, ArXiv:1703.07737.
- [10] ZHU Y, YANG Z, WANG L, et al. Hetero-center loss for cross-modality person re-identification [J]. Neurocomputing, 2020, 386: 97-109.
- [11] NGUYEN D T, HONG H G, KIM K W, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. Sensors, 2017, 17(3): 605.
- [12] YE M, LAN X, LI J, et al. Hierarchical discriminative learning for visible thermal person re-identification[C]. Thirty-Second AAAI Conference on Artificial Intelligence, 2018: 7501-7508.
- [13] HAO Y, WANG N, LI J, et al. HSME: Hypersphere manifold embedding for visible thermal person re-identification[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 8385-8392.
- [14] WANG Z, WANG Z, ZHENG Y, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 618-626.
- [15] WANG G A, ZHANG T, CHENG J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 3623-3632.
- [16] LIU H, CHENG J, WANG W, et al. Enhancing the discriminative feature learning for visible-thermal

- cross-modality person re-identification [J]. Neurocomputing, 2020, 398: 11-19.
- [17] LI D, WEI X, HONG X, et al. Infrared-visible cross-modal person re-identification with an X modality[C]. Proc. AAAI Conf. Artif. Intell., 2020: 4610-4617.
- [18] YE M, LAN X, LENG Q, et al. Cross-modality person re-identification via modality-aware collaborative ensemble learning[J]. IEEE Transactions on Image Processing, 2020, 29: 9387-9399.
- [19] YE M, SHEN J, CRANDALL D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification [J]. ArXiv Preprint, 2020, ArXiv:2007.09314.
- [20] WANG P, SU F, ZHAO Z, et al. Deep hard modality alignment for visible thermal person re-identification[J]. Pattern Recognition Letters, 2020, 133: 195-201.
- [21] LIU H, CHAI Y, TAN X, et al. Strong but simple baseline with dual-granularity triplet loss for visible-thermal person re-identification [J]. IEEE Signal Processing Letters, 2021, 28: 653-657.
- [22] WANG G A, ZHANG T, YANG Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12144-12151.
- [23] FENG Z, LAI J, XIE X. Learning modality-specific representations for visible-infrared person re-identification [J]. IEEE Transactions on Image Processing, 2019, 29: 579-590.

作者简介

周航, 硕士研究生, 主要研究方向为计算机视觉、模式识别、图像处理。

E-mail: 524563281@qq.com

黄春光, 讲师, 主要研究方向为混沌系统、随机数发生器。

E-mail: 2002041@hlju.edu.cn

程海(通信作者), 副教授, 硕导, 博士, 主要研究方向为混沌系统、格密码学、网络安全。

E-mail: chengh@hlju.edu.cn